A Unified Variational Approach to data analysis
Dynamic programming procedure

Tula State Technical University

Laboratory of Data Analysis

Members

Vadim V. Mottl	Head of the laboratory, D.Sci., Professor
Sergey D. Dvoyenko	Ph.D., Associate Professor
Andrey V. Kopylov	Ph.D., Associate Professor
Alexey A. Kostin	Ph.D. student
Oleg S. Seredin	Ph.D. student
Alexey S. Yermakov	Ph.D. student
Olga V. Krasotkina	Ph.D. student

A unified variational approach to data analysis.

Source data array

$$Y = (y_t, t \in T), y \in \mathcal{Y}$$

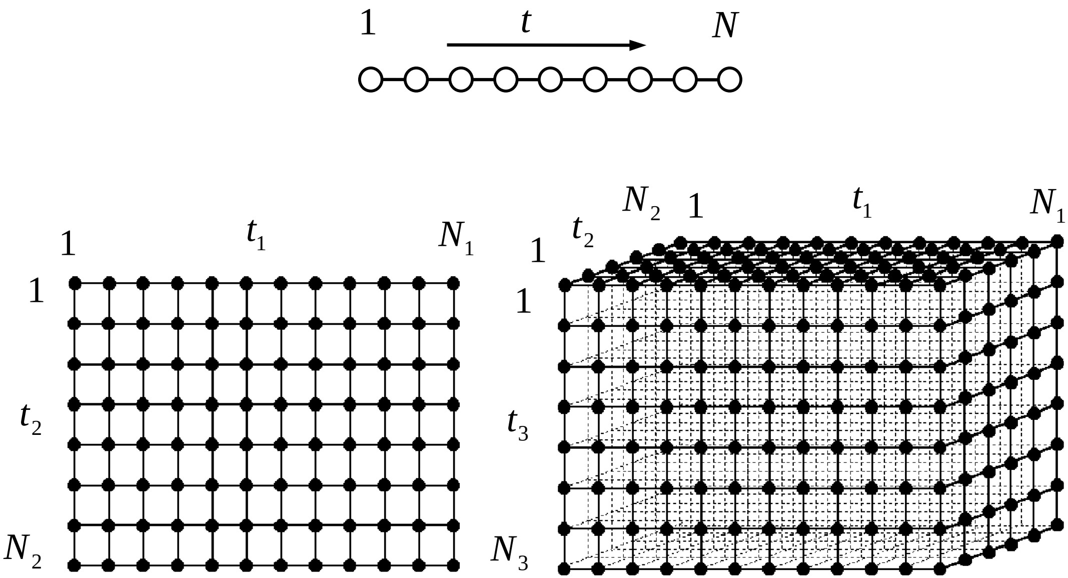
Result of processing

$$X = (x_t, t \in T), \; x \in \mathcal{X}$$

The original problem of data analysis can be turned to that of minimization of an appropriate objective function $J(X \mid Y)$ defined on the whole variety of possible results:

$$\hat{X}(Y) = \arg\min_{x_t \in \mathcal{X}} J(X \mid Y)$$

Adjacency graphs on the set of data array elements



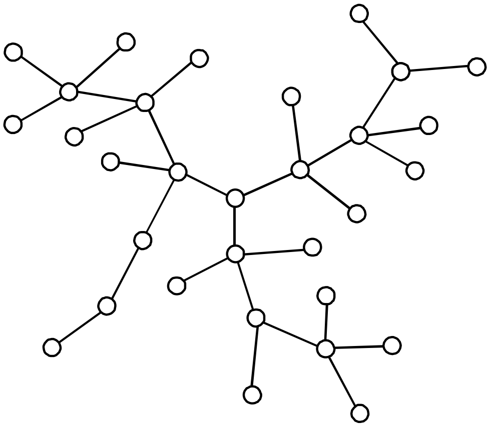
Pairwise separable objective function

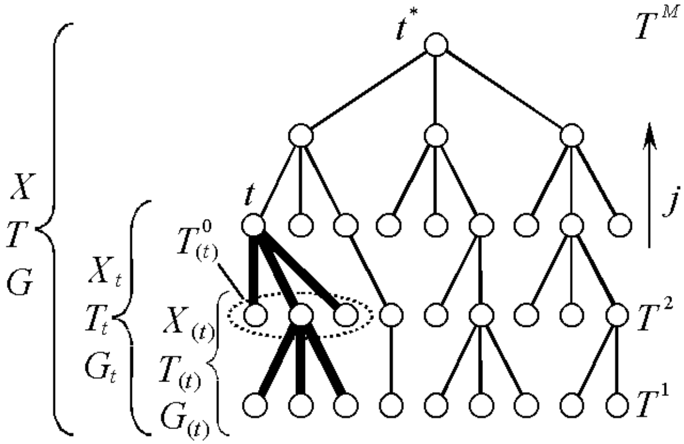
$$J(X\,|\,Y)=\sum_{t\in T}\psi_t(x_t\,|\,Y_t)+\sum_{(t',t'')\in G}\gamma_{t't''}(x_{t'},x_{t''})$$

$\psi_t(x_t\,|\,Y_t)$ – *node functions* which are meant to take the greater value the more evident is the contradiction between the hypothesis that x_t is just the “correct” local value we are seeking and the correspondent fragment Y_t of the source data array,

$\gamma_{t't''}(x_{t'},x_{t''})$ – *edge functions* that take the greater values the greater is the discrepancy between their arguments.

Hierarchical representation of a tree-like adjacency graph





Partial objective function defined on the descendant tree of node t

$$J_t(X_t) = \sum_{s \in T_t} \psi_s(x_s) + \sum_{(s',s'') \in G_t} \gamma_{s's''}(x_{s'}, x_{s''})$$

$$J_t(X_t) = J_t(x_t, X_{(t)}) = \psi_t(x_t) + \sum_{s \in T^0_{(t)}} \gamma_{ts}(x_t, x_s) + \sum_{s \in T^0_{(t)}} J_s(X_s) = \psi_t(x_t) + \sum_{s \in T^0_{(t)}} \left\{ \gamma_{ts}(x_t, x_s) + J_s(X_s) \right\}$$

Bellman functions

$$\tilde{J}_t(x_t) = \min_{X_{(t)}} J_t(x_t, X_{(t)}) = \psi_t(x_t) + \min_{x_s, X_{(s)}, s \in T^0_{(t)}} \sum_{s \in T^0_{(t)}} \left\{ \gamma_{ts}(x_t, x_s) + J_s(x_s, X_{(s)}) \right\}$$

Fundamental property of Bellman functions (upward recurrent relation)

$$\tilde{J}_t(x_t) = \psi_t(x_t) + \sum_{s \in T^0_{(t)}} \min_{x_s} \left\{ \gamma_{ts}(x_t, x_s) + \tilde{J}_s(x_s) \right\}$$

Downward recurrent relations

$$\tilde{x}_s(x_t) = \arg\min_{x_s} \left\{ \gamma_{ts}(x_t, x_s) + \tilde{J}_s(x_s) \right\}, \; s \in T^0_{(t)}$$

Marginal node functions

$$\hat{J}_t(x_t) = \min_{x_s,\, s \in T,\, s \neq t} J(X)$$

$$\hat{J}_s(x_s) = \min_{x_t} \left\{ \hat{J}_t(x_t) + \left[\tilde{J}_s(x_s) + \gamma_{ts}(x_t, x_s) \right] - \min_{x'_s} \left[\tilde{J}_s(x'_s) + \gamma_{ts}(x_t, x'_s) \right] \right\}, \; s \in T^0_{(t)}$$

Algorithm of objective function minimization

Ascending pass:

calculate and remember all the Bellman functions starting with Bellman functions for leaves $\tilde{J}_t(x_t) = \psi_t(x_t)$.

$$\tilde{J}_t(x_t) = \psi_t(x_t) + \sum_{s \in T^0_{(t)}} \min_{x_s} \left\{ \gamma_{ts}(x_t, x_s) + \tilde{J}_s(x_s) \right\}$$

Descending pass:

calculate optimal variable values or marginal functions for all the nodes starting with the root, where the marginal function coincides with the Bellman function $\hat{J}_{t^*}(x_{t^*}) = \tilde{J}_{t^*}(x_{t^*})$ and its minimization gives the optimal value at the root

$\hat{x}_{t^*} = \arg\min_{x_{t^*}} \tilde{J}_{t^*}(x_{t^*})$

$$\hat{J}_s(x_s) = \min_{x_t} \left\{ \hat{J}_t(x_t) + \left[\tilde{J}_s(x_s) + \gamma_{ts}(x_t, x_s) \right] - \min_{x'_s} \left[\tilde{J}_s(x'_s) + \gamma_{ts}(x_t, x'_s) \right] \right\},\ s \in T^0_{(t)}$$

Parametric family of quadratic functions for the case of continuous variables

Source data array	$Y=(y_t,\, t\in T),\, y\in \mathscr{Y}$
Result of processing	$X=(\mathbf{x}_t,\, t\in T),\ \mathbf{x}_t\in \mathrm{R}^n$
Quadratic node functions	$\psi_t(\mathbf{x}_t)=b_t^0+(\mathbf{x}_t-\mathbf{x}_t^0)^T\mathbf{Q}_t^0(\mathbf{x}_t-\mathbf{x}_t^0)$
Quadratic edge functions	$\gamma_{t't''}(\mathbf{x}_{t'},\mathbf{x}_{t''})=(\mathbf{x}_{t'}-\mathbf{x}_{t''})^T\mathbf{U}_{t't''}(\mathbf{x}_{t'}-\mathbf{x}_{t''})$
Pairwise separable quadratic objective function	$J(X)=\sum_{t\in T}\Big(b_t^0+(\mathbf{x}_t-\mathbf{x}_t^0)^T\mathbf{Q}_t^0(\mathbf{x}_t-\mathbf{x}_t^0)\Big)+\sum_{(t',t'')\in G}(\mathbf{x}_{t'}-\mathbf{x}_{t''})^T\mathbf{U}_{t't''}(\mathbf{x}_{t'}-\mathbf{x}_{t''})$

Parametric dynamic programming procedure for quadratic pairwise separable objective functions

Ascending pass: Bellman functions $\tilde{J}_t(\mathbf{x}_t)=\tilde{b}_t+(\mathbf{x}_t-\tilde{\mathbf{x}}_t)^T\tilde{\mathbf{Q}}_t(\mathbf{x}_t-\tilde{\mathbf{x}}_t)$ are calculated.

$$\tilde{\mathbf{Q}}_t=\mathbf{Q}_t^0+\sum_{s\in T_{(t)}^0}\tilde{\mathbf{Q}}_s\left(\tilde{\mathbf{Q}}_s+\mathbf{U}_{ts}\right)^{-1}\mathbf{U}_{ts},$$

$$\tilde{\mathbf{x}}_t=\tilde{\mathbf{Q}}_t^{-1}\left\{\mathbf{Q}_t^0\mathbf{x}_t^0+\sum_{s\in T_{(t)}^0}\tilde{\mathbf{Q}}_s\left(\tilde{\mathbf{Q}}_s+\mathbf{U}_{ts}\right)^{-1}\mathbf{U}_{ts}\tilde{\mathbf{x}}_s\right\},$$

$$\tilde{b}_t=b_t^0+(\mathbf{x}_t^0-\tilde{\mathbf{x}}_t)^T\mathbf{Q}_t^0\mathbf{x}_t^0+\sum_{s\in T_{(t)}^0}\left\{\tilde{b}_s+\left(\mathbf{x}_t-\tilde{\mathbf{x}}_s\right)^T\tilde{\mathbf{Q}}_s\left(\tilde{\mathbf{Q}}_s+\mathbf{U}_{ts}\right)^{-1}\mathbf{U}_{ts}\tilde{\mathbf{x}}_s\right\},$$

starting with the leaves, where $\tilde{J}_t(X_t)=\psi_t(x_t)$ and, consequently, $\tilde{\mathbf{x}}_t=\mathbf{x}_t^0$, $\tilde{b}_t=b_t^0$, $\tilde{\mathbf{Q}}_t=\mathbf{Q}_t^0$

Descending pass: marginal node functions $\hat{J}_t(\mathbf{x}_t)=\hat{b}+(\mathbf{x}_t-\hat{\mathbf{x}}_t)^T\hat{\mathbf{Q}}_t(\mathbf{x}_t-\hat{\mathbf{x}}_t)$ are calculated

$$\hat{b}=\tilde{b}_{t^*}$$

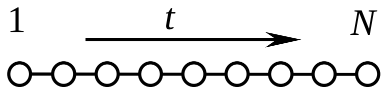
$$\hat{\mathbf{x}}_s=\tilde{\mathbf{x}}_s(\mathbf{x}_t)=(\mathbf{I}-\tilde{\mathbf{H}}_s)\tilde{\mathbf{x}}_s+\tilde{\mathbf{H}}_s\mathbf{x}_t=\tilde{\mathbf{x}}_t+\tilde{\mathbf{H}}_s(\mathbf{x}_t-\tilde{\mathbf{x}}_s),$$

$$\hat{\mathbf{Q}}_s=[\tilde{\mathbf{H}}_s\hat{\mathbf{Q}}_t^{-1}\tilde{\mathbf{H}}_s^T+(\tilde{\mathbf{Q}}_s+\mathbf{U}_{ts})^{-1}]^{-\mathbf{1}},$$

$$\tilde{\mathbf{H}}_s=(\tilde{\mathbf{Q}}_s+\mathbf{U}_{ts})^{-\mathbf{1}}\mathbf{U}_{ts}=(\mathbf{U}_{ts}^{-1}\tilde{\mathbf{Q}}_s+\mathbf{I})^{-1},\ s\in T_{(t)}^0,$$

starting from the root, where $\hat{J}_{t^*}(\mathbf{x}_{t^*})=\tilde{J}_{t^*}(\mathbf{x}_{t^*})$ and, hence, $\hat{\mathbf{x}}_{t^*}=\tilde{\mathbf{x}}_{t^*}$, $\hat{\mathbf{Q}}_{t^*}=\tilde{\mathbf{Q}}_{t^*}$.

Parametric dynamic programming procedure "forward and back" for signal analysis



Node functions $\psi_t(\mathbf{x}_t) = b_t^0 + (\mathbf{x}_t - \mathbf{x}_t^0)^T \mathbf{Q}_t^0 (\mathbf{x}_t - \mathbf{x}_t^0)$

Edge functions $\gamma_t(\mathbf{x}_{t-1}, \mathbf{x}_t) = (\mathbf{x}_{t-1} - \mathbf{A}\mathbf{x}_t)^T \mathbf{U}_t (\mathbf{x}_{t-1} - \mathbf{A}\mathbf{x}_t)$

Pairwise separable quadratic objective function $J(X) = \sum_{t=1}^{N} \Big(b_t^0 + (\mathbf{x}_t - \mathbf{x}_t^0)^T \mathbf{Q}_t^0 (\mathbf{x}_t - \mathbf{x}_t^0) \Big) + \sum_{t=2}^{N} \Big((\mathbf{x}_{t-1} - \mathbf{A}\mathbf{x}_t)^T \mathbf{U}_t (\mathbf{x}_{t-1} - \mathbf{A}\mathbf{x}_t) \Big)$

First (forward) pass along the signal $t = 1, \ldots, N$: parameters of Bellman functions are calculated

$$\tilde{\mathbf{Q}}_t = \mathbf{Q}_t^0 + \mathbf{A}^T \left(\tilde{\mathbf{Q}}_{t-1}^{-1} + \mathbf{U}_t^{-1} \right)^{-1} \mathbf{A},$$

$$\tilde{\mathbf{x}}_t = \tilde{\mathbf{Q}}_t^{-1} \left\{ \mathbf{Q}_t^0 \mathbf{x}_t^0 + \mathbf{A}^T \left(\tilde{\mathbf{Q}}_{t-1}^{-1} + \mathbf{U}_t^{-1} \right)^{-1} \tilde{\mathbf{x}}_{t-1} \right\},$$

$$\tilde{b}_t = b_t^0 + \tilde{b}_{t-1} + (\mathbf{x}_t^0 - \tilde{\mathbf{x}}_t)^T \mathbf{Q}_t^0 \mathbf{x}_t^0 + \left(\tilde{\mathbf{x}}_{t-1} - \mathbf{A}\mathbf{x}_t \right)^T \left(\tilde{\mathbf{Q}}_{t-1}^{-1} + \mathbf{U}_t^{-1} \right)^{-1} \tilde{\mathbf{x}}_{t-1},$$

the first sample is the only leaf $t = 1$: $\tilde{b}_1 = b_1^0$, $\tilde{\mathbf{x}}_1 = \mathbf{x}_1^0$, $\tilde{\mathbf{Q}}_1 = \mathbf{Q}_1^0$.

Second (backward) pass $t = N, \ldots, 1$: parameters of marginal node functions are computed

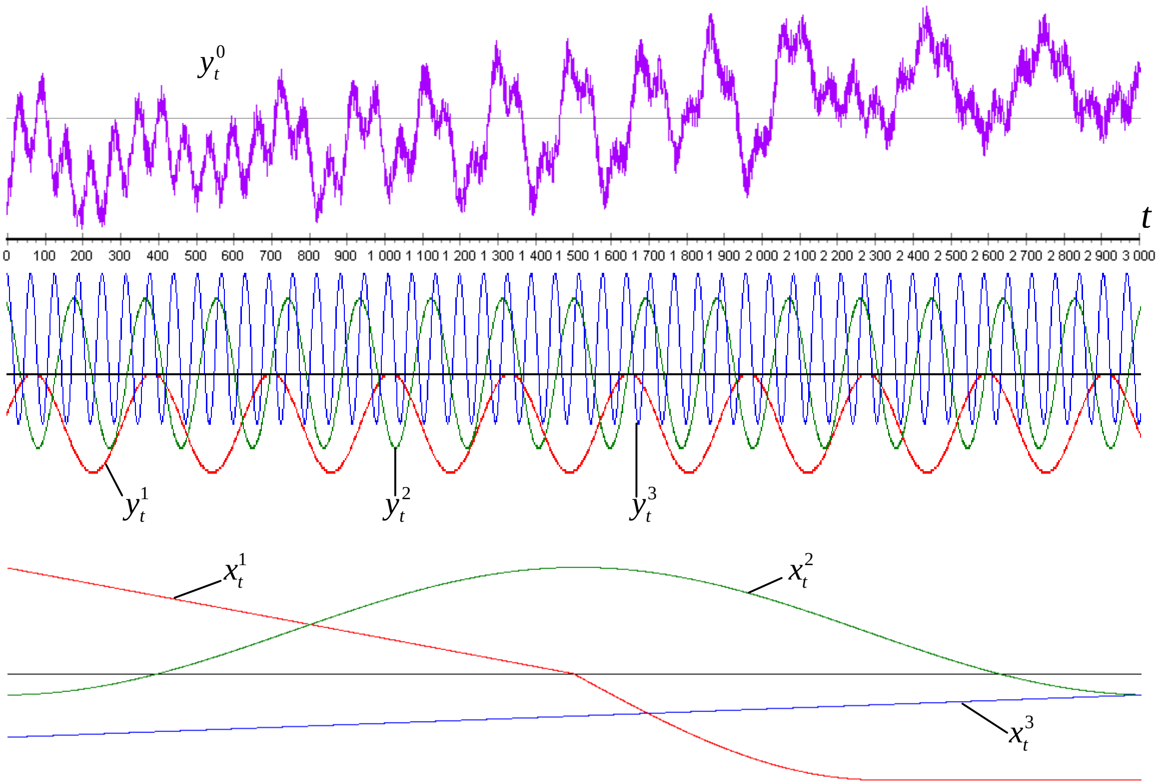
$$\hat{\mathbf{x}}_{t-1} = (\mathbf{I} - \tilde{\mathbf{H}}_{t-1}) \tilde{\mathbf{x}}_{t-1} + \tilde{\mathbf{H}}_{t-1} \mathbf{A}\mathbf{x}_t = \tilde{\mathbf{x}}_{t-1} + \tilde{\mathbf{H}}_{t-1} (\mathbf{A}\mathbf{x}_t - \tilde{\mathbf{x}}_{t-1})$$

$$\hat{\mathbf{Q}}_{t-1} = \left[\tilde{\mathbf{H}}_{t-1} \mathbf{A}_t \hat{\mathbf{Q}}_t^{-1} \mathbf{A}_t^T \tilde{\mathbf{H}}_{t-1}^T + (\tilde{\mathbf{Q}}_{t-1} + \mathbf{U}_t)^{-1} \right]^{-1}, \qquad \tilde{\mathbf{H}}_{t-1} = (\tilde{\mathbf{Q}}_{t-1} + \mathbf{U}_t)^{-1} \mathbf{U}_t$$

starting with the last sample of the signal $t = N$, where $\hat{\mathbf{x}}_N = \tilde{\mathbf{x}}_N$, $\hat{\mathbf{Q}}_N = \tilde{\mathbf{Q}}_N$

Estimation of non-stationary regression coefficients as the universal problem of signal processing

$$y_t^0 = \mathbf{x}_t^T \mathbf{y}_t + \xi_t$$



Estimation of non-stationary regression coefficients in terms of the proposed optimization procedure

Signal to be processed	$(y_t^0, \mathbf{y}_t),\ y_t^0 \in \mathrm{R},\ \mathbf{y}_t \in \mathrm{R}^n,\ t=1,...,N$
Sought-for sequence of non-stationary regression coefficients	$\mathbf{x}_t \in \mathrm{R}^n$
Model of non-stationary linear regression	$y_t^0 = \mathbf{x}_t^T \mathbf{y}_t + \xi_t$

Local quadratic criterion for estimating the current vector of non-stationary regression coefficients (node functions):

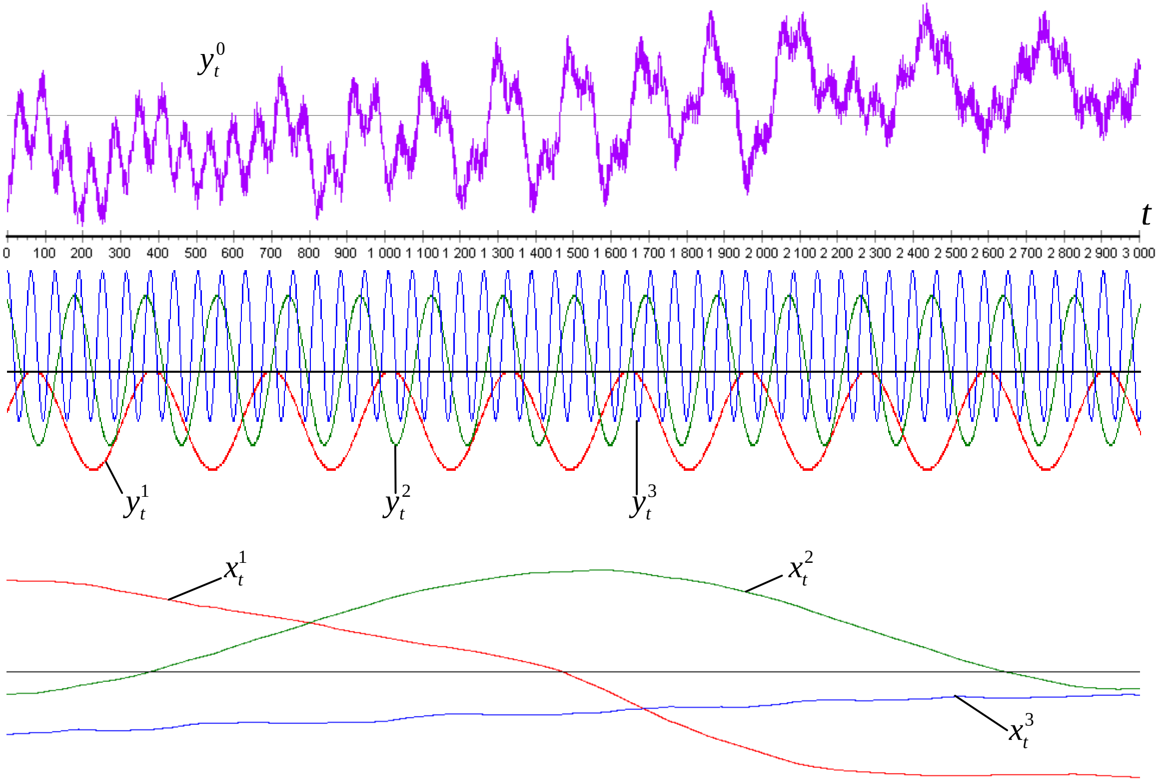
$$\psi_t(\mathbf{x}_t) = (y_t^0 - \mathbf{x}_t^T \mathbf{y}_t)^2 = (\mathbf{x}_t - \mathbf{x}_t^0)^T \mathbf{Q}_t^0 (\mathbf{x}_t - \mathbf{x}_t^0), \qquad \mathbf{x}_t^0 = \frac{y_t^0}{\mathbf{y}_t^T \mathbf{y}_t} \mathbf{y}_t, \qquad \mathbf{Q}_t^0 = \mathbf{y}_t \mathbf{y}_t^T .$$

Local quadratic penalty on unsmoothness of the sequence of non-stationary regression coefficient (edge functions):

$$\gamma_t(\mathbf{x}_{t-1}, \mathbf{x}_t) = (\mathbf{x}_{t-1} - \mathbf{x}_t)^T \mathbf{U}_t (\mathbf{x}_{t-1} - \mathbf{x}_t),$$

$\mathbf{U}_t$ – an appropriate sequence of positive definite matrices $n \times n$, which preset the desired degree of smoothing at each pair of neighboring samples $(t-1,\ t)$

An example of estimating non-stationary regression coefficients



Smoothing of a signal with estimation of its differences

Model of the source signal:

$$y_t = \mathbf{x}_t^T \mathbf{y}_t + \xi_t$$

$\mathbf{x}_t = (x_t^1, ..., x_t^n)^T \in \mathrm{R}^n$ – vector composed from the smoothed signal itself and its first $n-1$ differences

$$x_t^1 = x_t$$

$$x_t^2 = \nabla^1 x_t = x_t^1 - x_{t-1}^1$$

$$\ldots$$

$$x_t^n = \nabla^{n-1} x_t = x_t^{n-1} - x_{t-1}^{n-1}$$

$\mathbf{y}_t = (1,0,...,0)^T \in \mathrm{R}^n$ – vector whose first component is unit and all the others are zeros

Model of the sought-for signal

$$x_{t-1}^1 = x_t^1 - x_t^2$$

$$x_{t-1}^2 = x_t^2 - x_t^3$$

$$\ldots$$

$$x_{t-1}^n = x_t^n$$

$$\mathbf{x}_{t-1} = \mathbf{A}\mathbf{x}_t, \quad \mathbf{A} = \begin{pmatrix} 1 & -1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Quadratic node functions:

$$\psi_t(\mathbf{x}_t) = (y_t - x_t^0)^2 = (\mathbf{x}_t - \mathbf{x}_t^0)^T \mathbf{Q}_t^0 (\mathbf{x}_t - \mathbf{x}_t^0)$$

$$\mathbf{Q}_t^0 = \mathbf{y}_t \mathbf{y}_t^T = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \qquad \mathbf{x}_t^0 = \frac{y_t}{\mathbf{y}_t^T \mathbf{y}_t} \mathbf{y}_t = \begin{pmatrix} y_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

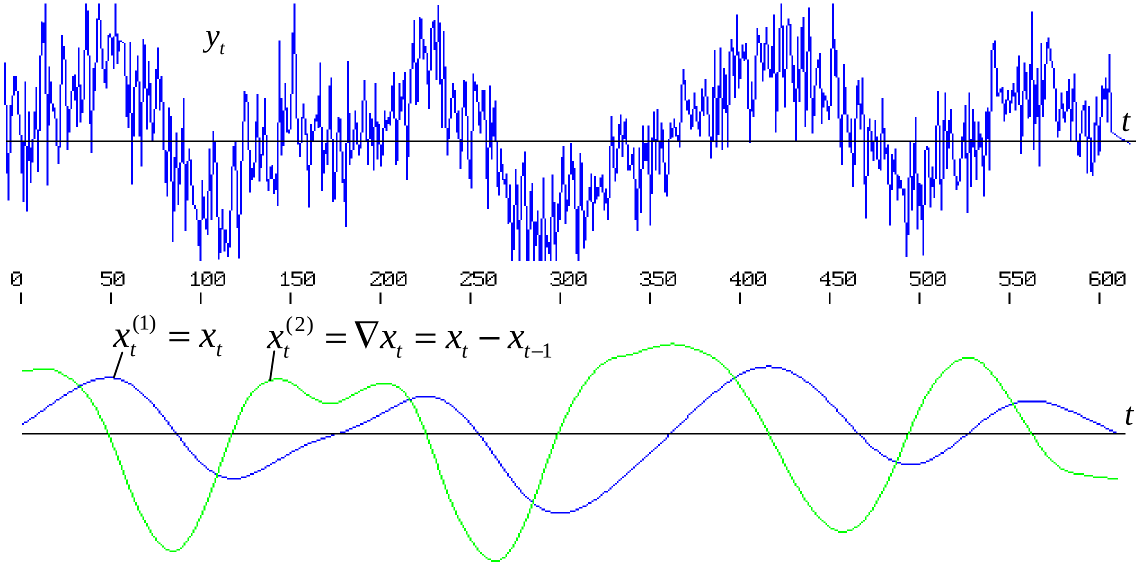
Quadratic edge function

$$\gamma_t(\mathbf{x}_{t-1}, \mathbf{x}_t) = (\mathbf{x}_{t-1} - \mathbf{A}\mathbf{x}_t)^T \mathbf{U}_t (\mathbf{x}_{t-1} - \mathbf{A}\mathbf{x}_t),$$

$$\mathbf{U}_t = \begin{pmatrix} \beta u_t & 0 & \cdots & 0 \\ 0 & \beta u_t & \vdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_t \end{pmatrix}$$

where β is a large positive number $\beta \to \infty$.

Example of smoothing a signal and its first difference



Autoregression and time-frequency analysis

Source signal

$$y_t = \mathbf{a}_t^T \mathbf{y}_t + \xi_t$$

$\mathbf{y}_t = (y_{t-1} \cdots y_{t-n})^T$ – prehistory vector composed from n previous samples of the source signal

$\mathbf{a} = (a_1 \cdots a_n)^T \in \mathrm{R}^n$ – vector of autoregression coefficients

Sought-for vector signal

$$\mathbf{x}_t = (\mathbf{a}_t, D_t) \in \mathrm{R}^{n+1}$$

Quadratic node functions

$$\psi_t(\mathbf{a}_t, D_t) = \psi_t(\mathbf{x}_t) = (\mathbf{x}_t - \mathbf{x}_t^0)^T \mathbf{Q}_t^0 (\mathbf{x}_t - \mathbf{x}_t^0),$$

$$\mathbf{x}_t^0 = \begin{pmatrix} \mathbf{a}_t^0 \\ D_t^0 \end{pmatrix} = \begin{pmatrix} \dfrac{y_t}{\mathbf{y}_t^T \mathbf{y}_t} \mathbf{y}_t \\ y_t^2 \end{pmatrix} \in \mathrm{R}^{n+1}, \quad \mathbf{Q}_t^0 = \begin{pmatrix} \mathbf{y}_t \mathbf{y}_t^T & \mathbf{0} \\ \mathbf{0}^T & 1 \end{pmatrix} [(n+1)\times(n+1)]$$

Quadratic edge functions

$$\gamma(\mathbf{x}_{t-1}, \mathbf{x}_t) = (\mathbf{x}_{t-1} - \mathbf{x}_t)^T \mathbf{U}_t (\mathbf{x}_{t-1} - \mathbf{x}_t),$$

$$\mathbf{U}_t = \begin{pmatrix} u_t \mathbf{I} & \mathbf{0} \\ \mathbf{0}^T & \alpha u_t \end{pmatrix},$$

$\mathbf{I}$ is unit matrix $(n \times n)$, the coefficient $\alpha > 0$ is meant to coordinate the smoothing degrees for signal intensity D and autoregression parameters $\mathbf{a}_t$.

Spectral representation of autoregression parameters

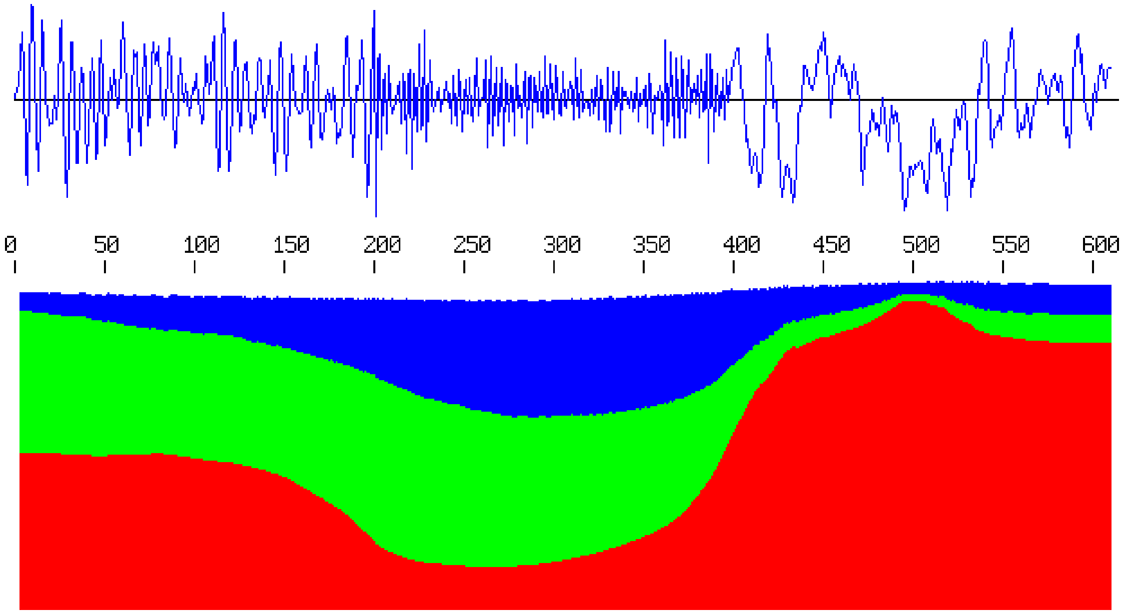
$$S(f)=\frac{(1+a_2)\left((1-a_2)^2-a_1^2\right)}{(1-a_2)}\frac{2D}{1+a_1^2+a_2^2-2a_1(1-a_2)\cos 2\pi f-2a_2\cos 4\pi f}$$

$$\hat{D}_t(0,1/6)=\int_0^{1/6}S(f;\hat{\mathbf{a}}_t,\hat{D}_t)df,$$

$$\hat{D}_t(1/6,1/3)=\int_{1/6}^{1/3}S(f;\hat{\mathbf{a}}_t,\hat{D}_t)df,$$

$$\hat{D}_t(1/3,1/2)=\int_{1/3}^{1/2}S(f;\hat{\mathbf{a}}_t,\hat{D}_t)df.$$

Result of processing a simulated signal



low, medium, and high frequency bands